

# SFedHIFI: Fire Rate-Based Heterogeneous Information Fusion for Spiking Federated Learning

Ran Tao<sup>1, 2</sup>, Qiugang Zhan<sup>1, 2</sup>, Shantian Yang<sup>1, 2, 3</sup>, Xiurui Xie<sup>4</sup>, Qi Tian<sup>5, 1</sup>, Guisong Liu<sup>1, 2, 3\*</sup>

<sup>1</sup>Complex Laboratory of New Finance and Economics, School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, China

<sup>2</sup>Engineering Research Center of Intelligent Finance, Ministry of Education, China

<sup>3</sup>Kash Institute of Electronics and Information Industry, China

<sup>4</sup>Laboratory of Intelligent Collaborative Computing, University of Electronic Science and Technology of China, China

<sup>5</sup>Huawei Inc., China

rantaostd@gmail.com, {zhanqg, yangst, gliu}@swufe.edu.cn, xiexiurui@uestc.edu.cn, tian.qi1@huawei.com

## Abstract

Spiking Federated Learning (SFL) has been widely studied with the energy efficiency of Spiking Neural Networks (SNNs). However, existing SFL methods require model homogeneity and assume all clients have sufficient computational resources, resulting in the exclusion of some resource-constrained clients. To address the prevalent system heterogeneity in real-world scenarios, enabling heterogeneous SFL systems that allow clients to adaptively deploy models of different scales based on their local resources is crucial. To this end, we introduce SFedHIFI, a novel Spiking Federated Learning framework with Fire Rate-Based Heterogeneous Information Fusion. Specifically, SFedHIFI employs channel-wise matrix decomposition to deploy SNN models of adaptive complexity on clients with heterogeneous resources. Building on this, the proposed heterogeneous information fusion module enables cross-scale aggregation among models of different widths, thereby enhancing the utilization of diverse local knowledge. Extensive experiments on three public benchmarks demonstrate that SFedHIFI can effectively enable heterogeneous SFL, consistently outperforming all three baseline methods. Compared with ANN-based FL, it achieves significant energy savings with only a marginal trade-off in accuracy.

## Introduction

Spiking Federated Learning (SFL) is an emerging research paradigm that integrates Spiking Neural Networks (SNNs) (Maass 1997) with Federated Learning (FL) (Nguyen et al. 2021; Huang et al. 2024), aiming to build FL systems tailored for resource-constrained edge environments. In particular, while FL serves as a general framework for privacy-preserving distributed learning, resource limitations of edge participants remain a critical challenge. Notably, SNNs employ an event-driven processing paradigm, enabling sparse and energy-efficient computation, especially when deployed on neuromorphic hardware, making them particularly suitable for resource-constrained environments (Merolla et al. 2014; Davies et al. 2018; DeBole

et al. 2019; Pei et al. 2019). Current SFL research primarily focuses on reducing the system power consumption by optimizing distributed training protocols and communication-efficient mechanisms (Skatchkovsky, Jang, and Simeone 2020; Venkatesha et al. 2021; Yang et al. 2022; Xie et al. 2024), demonstrating significant advancements in various common tasks (e.g., image classification).

Despite considerable progress in existing SFL, these approaches typically impose a strict requirement for architectural homogeneity between client models and the global one. This uniformity constraint proves overly restrictive for practical deployment scenarios. In contrast, participants typically utilize diverse edge devices capable of gathering valuable data, which are constrained by computational limitations and exhibit significant heterogeneity in resource budgets. This widespread system heterogeneity has emerged as a major bottleneck for the practical deployment of SFL applications in real-world scenarios.

To address system heterogeneity in FL, ANN-based approaches typically either exclude resource-constrained edge devices and their distinctive data knowledge (Xu and Wang 2020; Ji et al. 2023) or reduce resource demands through model compression techniques (Shah and Lau 2021; Xu et al. 2022), such as decreasing model depth (the number of layers) or width (the number of hidden channels). Motivated by the above strategies and recent advances in SNN pruning (Shi et al. 2019; Meng et al. 2023), we conducted preliminary explorations into a heterogeneous SFL system that combines channel pruning with conventional aggregation algorithms. However, as demonstrated by our empirical results in Table 1, the inherently sparse feature representations in SNNs are particularly susceptible to information loss during model compression, leading to degraded performance under heterogeneous settings. Hence, developing a heterogeneous SFL framework based on channel pruning is a promising direction. It can effectively train and aggregate information across diverse SNNs, yet remains underexplored.

In this paper, we propose SFedHIFI, a novel Spiking Federated Learning framework with Fire Rate-Based Heterogeneous Information Fusion. SFedHIFI supports the configuration of complex global SNN models and intro-

\*Corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

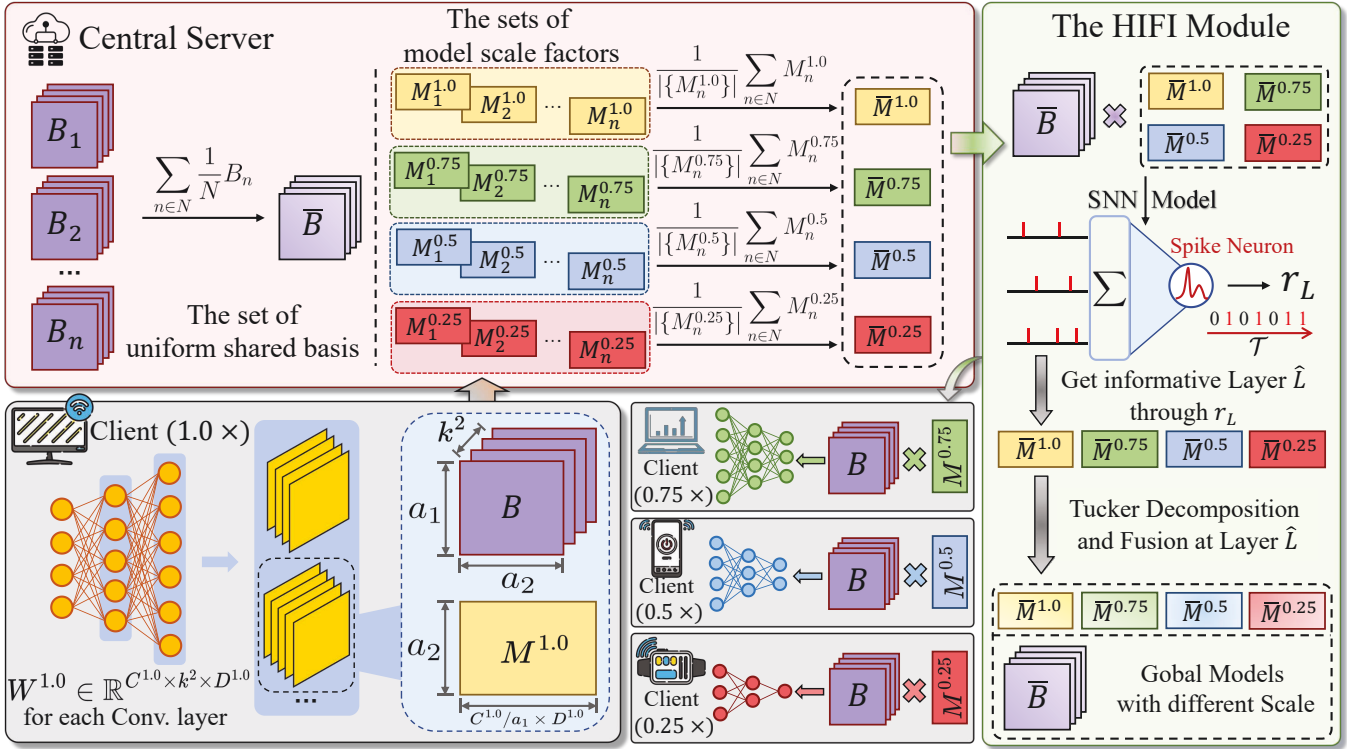


Figure 1: Overview of SFedHIFI. First, local clients are clustered into the model scale set of computational resources  $\mathbf{p} = \{0.25, 0.5, 0.75, 1.0\}$ . Then, the local ( $\mathbf{p} \times$ ) model gets the uniform shared basis  $B$  and the model scale factor  $M^{\mathbf{p}}$  through the Channel-wise Matrix Decomposition. In the central server, the uniform shared basis will be aggregated as the global  $\bar{B}$ , and the model scale factors will be aggregated into global  $\{\bar{M}^{\mathbf{p}}\}$  corresponding to different model scales. The list of global model scale factors  $\{\bar{M}^{\mathbf{p}}\}$  is then fed to the HIFI module, where the Firing Rate-Based Heterogeneous Information Fusion is applied to integrate knowledge across scales. This process enhances the representational capacity of global SNNs at all scales and improves their overall performance.

duces a dynamic channel pruning strategy to modify the width of local models for resource adaptation. Specifically, we introduce a channel-wise matrix decomposition technique (Astrid and Lee 2017; Phan et al. 2020) to project the parameters of SNNs with varying widths onto a shared  $B$  matrix representing the global channel space and a client-specific  $M$  matrix encoding heterogeneous channel dimensions. In this framework,  $B$  facilitates cross-scale aggregation from all models to capture broad knowledge, whereas  $M$  enables scale-specific aggregation of heterogeneous information, ensuring stable model convergence. Following aggregation, the matrices  $B$  and  $M$  are distributed to the corresponding clients. This approach guarantees universal access to distributed knowledge while significantly enhancing the representational diversity and generalization capacity of the global model.

Furthermore, due to the sparse and binary nature of SNN representations, the shared  $B$  matrix tends to capture limited global knowledge, while the client-specific matrices  $M$  retain potentially valuable cross-scale heterogeneous information. To leverage this, we propose a novel heterogeneous information fusion (HIFI) module tailored for the client-

specific  $M$  matrices to improve heterogeneous aggregation. Specifically, we analyze spiking neural dynamics to identify high-information-density layers within the  $M$  matrices, then decouple these valuable parameters and fuse them across different model scales.

Experimental results demonstrate that SFedHIFI, leveraging the spiking neural dynamics to address the information sparsity challenge inherent in SNNs, significantly enhances the performance of heterogeneous aggregation. The main contributions of our studies are summarized as follows:

- We propose SFedHIFI, a novel heterogeneous SFL framework that enables clients to adaptively deploy SNN models of varying widths via channel pruning, while effectively aggregating heterogeneous SNN information to improve global model performance across different scales.
- To address the challenge of severe information loss caused by SNN compression, we design a new HIFI module based on layer firing rates, which further exploits the intrinsic heterogeneous knowledge within SNNs and enhances the overall performance of SFedHIFI.
- Extensive experiments on four benchmark datasets

demonstrate that SFedHIFI effectively mitigates information degradation in compressed SNNs, achieving an energy-efficient heterogeneous SFL paradigm.

## Preliminary and Related Work

### The Leaky Integrate-and-Fire Model

In SNNs, information is encoded as binary spike sequences through spiking neural models, typically the Leaky Integrate-and-Fire (LIF) neurons (Gerstner et al. 2014; Doutsis et al. 2021):

$$H[t] = V[t - 1] - \tau(V[t - 1] - V_{\text{reset}}) + I[t], \quad (1)$$

$$S[t] = \Theta(V[t] - V_{\text{th}}), \quad (2)$$

$$V[t] = H[t] \cdot (1 - S[t]) + V_{\text{reset}} \cdot S[t], \quad (3)$$

where  $\tau$  is the leaky factor,  $V[t]$  represents the membrane potential of the neuron at time step  $t$ ,  $I[t]$  is the input for neuron.  $S[t]$  denotes the output state of a neuron, taking a value of 1 or 0 to indicate whether it is firing a spike or not.  $\Theta(\cdot)$  is the Heaviside function. When the membrane potential exceeds the threshold  $V_{\text{th}}$ , the neuron will trigger a spike and reset its membrane potential to  $V_{\text{reset}}$ .

### Spiking Federated Learning

Emerging research is exploring Spiking Federated Learning (SFL) to alleviate pervasive resource constraints at the edge in FL. Seminal work (Skatchkovsky, Jang, and Simeone 2020) pioneered on-device collaborative federated learning with SNNs. Tumpa et al. explored the integration of the Skipper technique (Singh et al. 2022) for optimizing SNN training protocols, effectively constraining training memory across clients (Tumpa et al. 2023). For communication efficiency in SFL, HDSFL leveraged hint-layer distillation (Xie et al. 2024), and Nguyen et al. introduced a Top-k sparsification strategy (Nguyen et al. 2024). Other studies focus on improving the effectiveness of SFL. For example, AdaFedAsy aimed to mitigate the negative effects of stale updates from resource-constrained devices (Wang, Duan, and Chen 2023), while Yu et al. explored the integration of higher-accuracy ANNs into the SFL framework (Yu et al. 2024). Collectively, these advancements substantiate the effectiveness of SFL and highlight its inherent advantages in energy efficiency.

Nevertheless, current SFL studies universally assume identical model architectures, making them unsuitable for the more common system heterogeneity scenarios. In this paper, we fill this gap and demonstrate the effectiveness of heterogeneous SNNs in federated learning, thereby broadening the applicability of SFL.

### System-Heterogeneous Federated Learning

To tackle system heterogeneity, SFL can draw inspiration from ANN-based FL studies. System heterogeneity primarily imposes resource constraints, which have been mitigated through accelerated convergence (Karimireddy et al. 2020), model compression (Haddadpour et al. 2021), and sparsified training (Qiu et al. 2022). However, these methods cause significant information loss in SNNs due to sparse binary

spikes. Moreover, FL based on knowledge distillation (KD) can support model heterogeneity, but its dependence on public datasets conflicts with privacy principles (Wu et al. 2022).

Other works leverage model pruning to assign client-specific subnetworks of different sizes, enabling adaptive knowledge contribution (Diao, Ding, and Tarokh 2021; Horvath et al. 2021; Mei et al. 2022; Jia et al. 2024). Given the effectiveness of SNN pruning (Shi et al. 2019; Meng et al. 2023), this provides valuable insight for heterogeneous SFL. Yet, these methods restrict aggregation to shared channels, which benefits dense ANNs but discards essential information in sparse SNNs, degrading performance (see Table 2). In contrast, our method estimates feature information density from neuron firing rates to guide the aggregation of heterogeneous SNNs, effectively improving the global model.

## Method

In this section, we first introduce the heterogeneous federated learning framework based on channel-wise matrix decomposition. Building upon this channel-pruned heterogeneous structure, we propose the HIFI module tailored for SNNs, which analyzes the firing rates across SNN layers to identify high-information-density layers within the client-specific  $M$  matrices. These informative layers are then decoupled using Tucker decomposition to enable effective heterogeneous information fusion. At last, we explain the local objective in SFedHIFI.

### Heterogeneous FL Framework Based on Channel-Wise Matrix Decomposition

Considering prior studies that have demonstrated the effectiveness of channel pruning in SNNs, our work adopts a channel-wise matrix decomposition strategy for model pruning, enabling the construction of a heterogeneous SFL framework that supports cross-scale model aggregation. Specifically, the heterogeneous FL framework comprises  $N$  local clients with models of varying scales and a central server. Every client  $n \in \{1, \dots, N\}$  has own local dataset. At the start, clients are clustered into the model scale set of computational resources  $\mathbf{p} = \{p_i | p_i \in (0, 1], i < N\}$ , according to their local computational resources. Next, the server assigns the uniform shared basis  $B$  and the model scale factor  $M^{\mathbf{p}}$  to each client according to the clustering results.

For a client  $n$ , its local model width is  $p_i \times$  full model, and the channel-wise matrix decomposition of the convolution weight parameter  $\mathbf{W}_n^{p_i}$  can be described as,

$$\begin{aligned} \mathbf{W}_n^{p_i} &\approx B_n \cdot M_n^{p_i}, \\ \mathbf{W}_n^{p_i} &\in \mathbb{R}^{C^{p_i} \times k^2 \times D^{p_i}}, \\ B_n &\in \mathbb{R}^{k^2 \times a_1 \times a_2}, M_n^{p_i} \in \mathbb{R}^{a_2 \times \frac{C^{p_i}}{a_1} \times D^{p_i}}, \end{aligned} \quad (4)$$

where  $C^{p_i}$  denotes the output channel with scale  $p_i$ ,  $D$  means the input channel, and  $k$  is the convolution kernel size.  $a_1$  and  $a_2$  are pre-defined hyperparameters used to construct the unified shared basis  $B$ .

Therefore, the shared basis  $B$  and the scale factor  $M^{p_i}$  are updated during every local training, then sent to the server for aggregation after local training. Specifically, the central

server aggregates  $B$  for all clients, and aggregates  $M^{p_i}$  for all clients with scale  $p_i$ . The global  $\bar{B}$  and  $\bar{M}^{p_i}$  can be expressed as follows:

$$\begin{aligned}\bar{B} &= \text{FA}(\{B_n\}_{n \in N}) \\ &= \sum_{n \in N} \frac{1}{N} B_n \in \mathbb{R}^{k^2 \times a_1 \times a_2},\end{aligned}\quad (5)$$

$$\begin{aligned}\bar{M}^{p_i} &= \text{FA}(\{M_n^p | p = p_i, n \in N\}) \\ &= \frac{1}{|\{M_n^p\}|} \sum_{n \in N} M_n^p \in \mathbb{R}^{a_2 \times \frac{c^{p_i}}{a_1} \times D^{p_i}},\end{aligned}\quad (6)$$

where  $B_n$  is the  $B$  matrix of client  $n$ , and  $\{M_n^p\}$  is the set of  $M$  with scale  $p_i$  from all clients.  $\text{FA}(\cdot)$  means the FedAVG algorithm (McMahan et al. 2017). After aggregation, the server would get the global  $\bar{B}$  and  $\{\bar{M}^{\mathbf{P}}\}$  list, and then distributes the set of global model weights  $\{\bar{\mathbf{W}}^{\mathbf{P}} = \bar{B} \cdot \bar{M}^{\mathbf{P}}\}$  for every scale in set  $\mathbf{p}$ .

However, experimental results reveal that SNNs perform suboptimally under the heterogeneous FL framework based solely on channel-wise matrix decomposition (as evidenced by the performance of HIFI<sub>w/o</sub> in Table 2 compared to the Baseline methods in Table 1). To address this limitation and enhance the effectiveness of SNNs in heterogeneous settings, we further propose the core component of the SFed-HIFI framework—the HIFI module.

### Firing Rate-Based Heterogeneous Information Fusion (HIFI) Module

In the above heterogeneous FL framework, the global  $\bar{B}$  measures information across all model scales, and the global list  $\{\bar{M}^{\mathbf{P}}\}$  contains heterogeneous information for each model scale. In heterogeneous settings, each model scale is typically associated with a limited number of clients and data, which can lead to severe information loss, particularly detrimental for SNNs with inherently sparse feature representations. On the other hand, the spike firing rate in SNNs serves as a distinctive property that effectively reflects the information density received by each network layer.

Thus, this section proposes a novel heterogeneous information fusion module based on firing rates, called the HIFI module. The key idea of HIFI is to identify the most informative layer parameters based on their firing rates and fuse these heterogeneous weights through information decoupling. This approach enriches the information representation of global SNNs and further enhances overall model performance. Firstly, we define the information density of layer  $L$  via the following layer spike firing rate  $r_L$ ,

$$r_L = \frac{1}{\mathcal{T} \cdot \|L\|} \sum_{t=1}^{\mathcal{T}} \sum_{l \in \text{layer } L} S_l[t], \quad (7)$$

where  $l$  means the spike neuron in Layer  $L$ , and  $\|L\|$  denotes the number of all spike neurons in Layer  $L$ ,  $S_l[t]$  is the output of neuron  $l$  at time-step  $t$ ,  $\mathcal{T}$  is the total time window length.

Then, we compare the firing rates  $r_L$  across all model scales in  $\mathbf{p}$  to identify the layer with the highest information

density  $\hat{L}$ , which is subsequently used for the heterogeneous information fusion,

$$\hat{L} = \text{mode}(\{L^{p_i} | p_i \in \mathbf{p}\}), \quad (8)$$

$$L^{p_i} = \underset{L}{\text{argmax}}(\{r_L^{p_i} | L \in \text{layer } 1, 2, \dots\}), \quad (9)$$

where  $\text{mode}(\cdot)$  means the mode extraction,  $r_L^{p_i}$  denotes the spike firing rate of layer  $L$  with  $p_i \times$  width. After identifying the layer with the highest information density, we perform heterogeneous information decoupling on the  $\hat{L}$  layer of each model scale, following the Tucker decomposition framework (Kolda and Bader 2009; Balažević, Allen, and Hospedales 2019),

$$\bar{M}_L^{p_i} \approx G^{p_i} \times_1 A_1^{p_i} \times_2 A_2^{p_i} \times_3 A_3^{p_i} \times_4 A_4^{p_i}, \quad (10)$$

where  $\bar{M}_L^{p_i}$  is the model scale factor of layer  $L$  in the global  $p_i \times$  model.  $G^{p_i} \in \mathbb{R}^{b_1 \times b_2 \times b_3}$  is the core tensor, and it retains the same shape when decomposed from all model scale factors  $\{\bar{M}_L^{p_i}\}$ .  $b_1, b_2$  and  $b_3$  are pre-defined Tucker ranks.  $[A_1^{p_i}, A_2^{p_i}, A_3^{p_i}, A_4^{p_i}]$  is the list of two-dimensional factor matrices, and  $\times_k$  is the Mode- $k$  Product for tensor and matrix. Finally, we aggregate the core tensors to get  $\bar{G}$  and reconstruct  $\{\bar{M}_L^{\mathbf{P}}\}$  using the factor matrices,

$$\bar{G} = \frac{1}{|\mathbf{P}|} \sum_{p_i \in \mathbf{P}} G^{p_i} \in \mathbb{R}^{b_1 \times b_2 \times b_3}, \quad (11)$$

$$\bar{M}_L^{\mathbf{P}} \leftarrow \bar{G} \times_1 A_1^{\mathbf{P}} \times_2 A_2^{\mathbf{P}} \times_3 A_3^{\mathbf{P}} \times_4 A_4^{\mathbf{P}}. \quad (12)$$

In summary, to mitigate the susceptibility of SNNs to information loss caused by their inherent sparse representations, the SFedHIFI framework identifies high-information-density layers via their spiking firing rates, then performs heterogeneous information fusion of global model scale factors  $\{\bar{M}^{\mathbf{P}}\}$  through information decoupling, thereby further enhancing the performance of the global model.

### The Local Objective Function

Our SFedHIFI employs the temporal efficient training (TET) loss (Deng et al. 2022), specifically adapted for SNNs, as the primary optimization objective for local training. As shown in Eq.4, the training process simultaneously optimizes both the parameter matrices  $B$  and  $M^{\mathbf{P}}$ . To enhance the global generalization capacity, the shared basis  $B$  needs to encode comprehensive knowledge representations. However, the standard objective function fails to effectively constrain the training dynamics of  $B$  to achieve this goal. To this end, we introduce an additional orthogonal regularizer into the local loss function, following (Mei et al. 2022),

$$\mathcal{L} = \mathcal{L}_{tet}(\mathbf{y}, y) + \lambda \cdot \sum_{L} \|B_L \cdot B_L^T - W_I\|_2, \quad (13)$$

$$\mathcal{L}_{tet} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \mathcal{L}_{CE}(\mathbf{y}[t], y), \quad (14)$$

where  $\lambda$  is the balancing factor, and  $W_I$  means identity matrix.  $\mathbf{y}[t]$  denotes the predicted output of SNN at time step  $t$ ,  $y$  is the label.  $B_L$  is the shared basis of layer  $L$ ,  $B_L^T$  is the transpose of  $B_L$ . In this work, we impose an orthogonal constraint on  $B$  via  $L_2$  normalization, which implicitly enriches its knowledge representation capacity.

Dataset	Scale	IID				Non-IID Dir(0.5)				Non-IID Dir(0.3)			
		FedAVG	FedProx	FedNova	SFedHIFI	FedAVG	FedProx	FedNova	SFedHIFI	FedAVG	FedProx	FedNova	SFedHIFI
Fashion-MNIST	0.25×	90.50	89.10	88.48	<b>91.46</b>	85.19	<b>89.21</b>	88.59	88.70	83.35	86.24	<b>87.74</b>	87.27
	0.5×	91.12	89.56	84.85	<b>92.38</b>	85.36	88.56	87.63	<b>89.85</b>	83.86	<b>88.70</b>	87.97	88.31
	0.75×	91.47	89.68	73.70	<b>92.03</b>	88.18	89.69	89.82	<b>90.41</b>	86.14	88.91	<b>89.31</b>	88.99
	1.0×	91.17	90.53	90.20	<b>92.43</b>	89.91	<b>90.79</b>	90.74	90.32	87.98	88.29	88.63	<b>89.18</b>
	Avg.	91.06	89.97	84.31	<b>92.08</b>	87.16	89.56	89.20	<b>89.82</b>	85.33	88.04	88.41	<b>88.43</b>
CIFAR-10	0.25×	85.52	84.96	85.43	<b>85.64</b>	78.49	78.53	79.57	<b>80.83</b>	54.68	52.06	47.82	<b>79.35</b>
	0.5×	87.26	89.58	88.78	<b>89.63</b>	85.21	85.41	85.62	<b>87.12</b>	79.60	76.98	74.25	<b>86.63</b>
	0.75×	<b>91.28</b>	91.04	90.06	91.14	88.71	88.76	87.97	<b>89.33</b>	81.22	72.71	68.26	<b>86.03</b>
	1.0×	92.02	91.52	90.32	<b>92.06</b>	87.72	87.90	88.37	<b>89.26</b>	82.46	77.75	75.24	<b>86.90</b>
	Avg.	89.02	89.28	88.65	<b>89.62</b>	85.03	85.15	85.38	<b>86.64</b>	74.49	69.88	66.39	<b>84.73</b>
CIFAR-100	0.25×	26.88	27.01	<b>27.25</b>	27.06	8.09	7.01	12.17	<b>20.53</b>	7.41	8.10	10.13	<b>19.40</b>
	0.5×	37.02	37.61	37.55	<b>37.96</b>	16.75	17.43	18.27	<b>31.80</b>	16.50	15.73	16.87	<b>29.46</b>
	0.75×	43.27	43.55	43.76	<b>43.94</b>	23.63	27.22	25.08	<b>39.66</b>	27.81	25.64	23.76	<b>39.37</b>
	1.0×	47.61	47.64	47.75	<b>47.84</b>	29.10	28.63	30.14	<b>40.01</b>	30.36	32.64	28.46	<b>39.72</b>
	Avg.	38.70	38.95	39.08	<b>39.20</b>	19.39	20.07	21.42	<b>33.00</b>	20.52	20.53	19.81	<b>31.99</b>

Table 1: Experimental results on Benchmark Comparison Evaluation. We report classification performance using the standard Top-1 accuracy (%) as the evaluation metric. The column labeled "Avg." denotes the average accuracy across all model scales, providing a comprehensive view of overall performance. Results highlighted in "numbers" indicate the best performance achieved within each comparison group.

## Experiments

This section provides a comprehensive evaluation of the proposed SFedHIFI framework. We first benchmark it against several classical FL methods with SNN, then perform ablation studies to assess the contribution of HIFI module. Finally, we compare SFedHIFI with ANN-based heterogeneous FL in terms of performance and energy efficiency. The source code and Appendix are available at <https://github.com/rtao499/SFedHIFI>.

### Experimental Setup

**Datasets.** We conducted experiments using three popular datasets, Fashion-MNIST (Xiao, Rasul, and Vollgraf 2017), CIFAR-10, and CIFAR-100 (Krizhevsky, Hinton et al. 2009). And like previous works (Li, He, and Song 2021; Xiang et al. 2024), we partition the datasets across clients under both IID and Non-IID data settings. The specifications of the dataset and Data Partitioning are detailed in the Appendix.

**Model Selection.** In terms of model structure, we adopt an identical SNN with the same setting for each dataset to ensure fair comparison. For Fashion-MNIST, we choose a simple 4-layer SNN (64C3-128C3-MP2-128C3-MP2-FC). And we use Spiking-ResNet18 provided by Spikejelly (Fang et al. 2023) for CIFAR-10 and CIFAR-100. For the SNN implementation, we set LIF neurons with threshold voltage  $V_{th} = 1.0$ , and the total time step  $\mathcal{T}$  is set to 10.

**Heterogeneity Settings.** In all experiments, we assume a total of 10 clients, with 50% of them randomly selected to participate in each communication round of SFedHIFI. To simulate system heterogeneity in model capacity, we set four distinct network scales, represented as  $\mathbf{p} = \{0.25, 0.5, 0.75, 1.0\}$ . For local data heterogeneity, we follow a Dirichlet-based partitioning strategy, setting the concentration parameter  $\alpha$  to 0.5 and 0.3, respectively, to reflect varying degrees of non-IID data distributions across clients.

**Baseline Approaches and Implementation Details.** For a comprehensive comparison, we compare our SFedHIFI

with three well-established FL baselines, including: FedAVG (McMahan et al. 2017), FedProx (Li et al. 2020), and FedNova (Wang et al. 2020). Due to model heterogeneity, the baseline methods perform aggregation only among models of the same scale during the FL process. Further details of baseline and implementation are available in the Appendix.

### Benchmark Comparison Evaluation

In the benchmark comparison experiments, we demonstrate the advantages of SFedHIFI by comparing against three baselines of FedAVG, FedProx, and FedNova adapted to SFL across three standard datasets. Specifically, we simulate an SFL environment with 10 clients, each assigned a local dataset following either an IID or non-IID distribution. In each communication round, 5 active clients are randomly selected to participate, each equipped with a heterogeneous SNN model. Moreover, we define four different SNN model scales based on network width to build model heterogeneity. While SFedHIFI performs cross-scale aggregation across all four model types, the baseline methods apply their respective aggregation strategies only among models of the same scale.

Table 1 shows that the proposed SFedHIFI consistently outperforms all baselines across all model scales, datasets, and data distributions. On the relatively simple Fashion-MNIST dataset, the performance gap among all methods is within 2.5% accuracy. Under the IID setting, FedAVG achieves the best performance among the baselines. In contrast, SFedHIFI benefits from the fusion of heterogeneous information across all model scales, allowing it to learn a broader knowledge base and achieve the best results at every model scale, with an average accuracy improvement of 1.02%. As the degree of non-IID increases, although SFedHIFI still maintains superior overall performance, FedProx and FedNova begin to demonstrate their advantages in handling data heterogeneity, outperforming FedAVG and performing comparably to SFedHIFI. These results not only

Scale	IID				Non-IID Dir(0.5)				Non-IID Dir(0.3)			
	HIFI <sub>w/o</sub>	HIFI <sub>fixed</sub>	HIFI <sub>rand</sub>	SFedHIFI	HIFI <sub>w/o</sub>	HIFI <sub>fixed</sub>	HIFI <sub>rand</sub>	SFedHIFI	HIFI <sub>w/o</sub>	HIFI <sub>fixed</sub>	HIFI <sub>rand</sub>	SFedHIFI
0.25×	83.82	82.11	77.72	<b>85.64</b>	80.65	74.25	73.60	<b>80.83</b>	74.52	68.94	58.57	<b>79.35</b>
0.5×	88.83	89.07	88.36	<b>89.63</b>	84.87	85.90	83.56	<b>87.12</b>	76.16	82.20	72.70	<b>86.63</b>
0.75×	90.55	91.06	89.96	<b>91.14</b>	85.55	88.94	86.61	<b>89.33</b>	82.68	82.26	79.19	<b>86.03</b>
1.0×	91.62	91.81	91.15	<b>92.06</b>	85.42	89.08	85.98	<b>89.26</b>	79.29	84.39	59.46	<b>86.90</b>
Avg.	88.71	88.52	86.80	<b>89.62</b>	84.12	84.54	82.44	<b>86.64</b>	78.16	79.45	67.48	<b>84.73</b>

Table 2: Ablation experiment results of the HIFI module on CIFAR-10. We still report classification performance using the standard Top-1 accuracy (%) as the evaluation metric.

validate the effectiveness of SFedHIFI in heterogeneous information fusion but also indicate that the simplicity of Fashion-MNIST may limit its ability to highlight the differences among methods.

On more complex datasets, SFedHIFI demonstrates more significant improvements. On CIFAR-10, the accuracy gains can reach up to 30%. Under the IID setting, all four methods perform similarly, as models of all scales have equal access to the same data knowledge. Nevertheless, SFedHIFI still achieves the best overall performance, showcasing the advantage of heterogeneous information fusion. As the degree of non-IID increases, the superiority of our method becomes more pronounced, especially under *Dir*(0.3). In this case, clients contain highly diverse data knowledge, and aggregating within homogeneous model structures not only limits knowledge sharing but also exacerbates performance degradation for models with a smaller scale. In contrast, SFedHIFI significantly improves the performance of small-scale models compared to baseline methods, confirming its effectiveness. These results also validate that cross-scale aggregation can successfully fuse heterogeneous information, enabling compact models to acquire deeper and broader knowledge.

On CIFAR-100, none of the methods achieved high absolute performance due to the dataset’s inherent complexity. However, this complexity also provides a clearer differentiation among methods. Similar to previous datasets, SFedHIFI shows a slight advantage under the IID setting. Under non-IID settings, baseline models at all scales struggle to learn effectively from local data. This highlights the inconsistency of data knowledge across clients, where aggregating only within the same model scale fails to build a comprehensive knowledge base. Consequently, large-scale models suffer from incomplete learning, while the limited capacity of small-scale models is further exposed without the guidance of larger models. In contrast, SFedHIFI consistently improves models of all scales, outperforming baselines by up to 19% in accuracy. These results demonstrate that, even on complex datasets, SFedHIFI can effectively expand the knowledge coverage of all clients by fusing heterogeneous information across models of different scales.

### Ablation Experiment of the HIFI Module

In the proposed SFedHIFI framework, we design a tailored HIFI Module for SNNs. To evaluate its effectiveness, we conduct experiments on the CIFAR-10 dataset. Specifically, the HIFI Module identifies high-information-density layers

based on spike firing rates and performs heterogeneous information fusion across different model scales. To analyze the impact of this design, we conduct three comparative experiments: (i) HIFI<sub>w/o</sub> – without heterogeneous information fusion; (ii) HIFI<sub>fixed</sub> – information fusion is performed on a fixed layer; and (iii) HIFI<sub>rand</sub> – information fusion is performed on randomly selected layers. The experimental results are presented in Table 2.

From ablation results in Table 2, our proposed method achieves the best performance across all comparisons, with particularly significant gains under the *Dir*(0.3) setting. Specifically, in all scenarios, HIFI<sub>rand</sub> performs the worst, indicating that the informativeness and error sensitivity of SNN layers vary significantly. Randomly selecting layers for aggregation may even introduce harmful effects.

When comparing SFedHIFI with HIFI<sub>w/o</sub>, we observe that under IID settings, the HIFI module brings more substantial improvements to smaller models. This is because, with uniform data distributions, large models gain little from small models in terms of their local data knowledge, whereas small models benefit from the parameter guidance of larger models. Under non-IID settings, SFedHIFI provides significant improvements across all model scales, confirming the effectiveness of the HIFI module. These results also validate that due to the sparsity of SNN representations and the information loss induced by pruning, removing the HIFI module results in limited aggregation effectiveness.

In the comparison between SFedHIFI and HIFI<sub>fixed</sub>, SFedHIFI yields greater improvements for smaller models. It is worth noting that the fixed layer in HIFI<sub>fixed</sub> is set to the last convolutional layer of Spiking-ResNet18, which is empirically regarded as one of the most informative layers in conventional ResNet18 architectures. However, our results suggest that information distribution in SNNs does not entirely align with ANN-based assumptions. By leveraging spike firing rates, our approach enables a more refined identification of informative layers, thereby capturing richer and more effective heterogeneous information for aggregation.

### Energy Analysis

SNNs have been widely recognized as a promising low-power alternative to ANNs. Therefore, we are particularly interested in comparing the advantages of ANN-based FL and SFL. To enable a more intuitive comparison between SNNs and ANNs, we evaluate both model performance and energy consumption on the Fashion-MNIST dataset. Specif-

Metric	Scale	IID		Dir(0.5)		Dir(0.3)	
		ANNs	Ours	ANNs	Ours	ANNs	Ours
Accuracy (%) ↑	0.25×	89.75	<b>91.46</b>	<b>90.91</b>	88.70	<b>90.62</b>	87.27
	0.5×	90.14	<b>92.38</b>	<b>91.74</b>	89.85	<b>91.00</b>	88.31
	0.75×	90.90	<b>92.03</b>	<b>91.33</b>	90.41	<b>91.04</b>	88.99
	1.0×	91.01	<b>92.43</b>	<b>91.90</b>	90.32	<b>91.56</b>	89.18
	Avg.	90.45	<b>92.08</b>	<b>91.47</b>	89.82	<b>91.06</b>	88.43
Energy (×10 <sup>-3</sup> mj) ↓		ANNs		Ours			
	0.25×	34.059		<b>23.181</b>			
	0.5×	69.503		<b>43.354</b>			
	0.75×	108.639		<b>62.320</b>			
	1.0×	151.469		<b>75.737</b>			
Avg.	90.917		<b>51.148</b>				

Table 3: Comparison between SFedHIFI and ANN in Accuracy and Energy Consumption under Identical Network Configurations on Fashion-MNIST.

ically, we replace the SNN models in SFedHIFI with ANNs of the same scale and remove the Heterogeneous Information Fusion module to construct a corresponding ANN-based FL system.

In terms of model performance, we compare the accuracy of global models at four different network scales. For energy consumption, we follow prior work (Pan et al. 2020) by estimating the synaptic operations (SynOPs) of SNNs using the accumulated number of multiply-and-accumulate (MAC) and spike-based accumulate (AC) operations. For ANNs, we compute energy consumption based on floating-point operations (FLOPs), following the method in (Wu et al. 2021). We provide the energy estimation formulas used for both SNN and ANN models below:

$$E_{\text{ANN}} = \sum_{l=1}^L (\text{FLOPs}^{(l)} \times E_{\text{MAC}}) \quad (15)$$

$$= \sum_{l=1}^L 2(f_{\text{in}}^l + 1)\beta^l \times E_{\text{MAC}},$$

$$E_{\text{SNN}} = (\text{FLOPs}^{(1)} \times E_{\text{MAC}}) + \sum_{l=2}^L (\text{SynOPs}^{(l)} \times E_{\text{AC}})$$

$$= (\text{FLOPs}^{(1)} \times E_{\text{MAC}}) + \sum_{l=2}^L \text{FLOPs}^{(l)} \times \bar{r}^{(l)} \times \mathcal{T} \times E_{\text{AC}}, \quad (16)$$

where  $L$  is the total number of layers in both ANN and SNNs, and  $l$  is the layer index,  $\text{FLOPs}^{(l)}/\text{SynOPs}^{(l)}$  is  $\text{FLOPs}/\text{SynOPs}$  in the  $l$ -th layer.  $f_{\text{in}}^l$  denotes the number of incoming connections to a neuron of the  $l$ -th layer.  $\beta^l$  represents the number of neurons in the  $l$ -th layer.  $\bar{r}^{(l)}$  means the average firing rate per time step of the neurons in the  $l$ -th layer.  $E_{\text{MAC}}$  is the energy cost per MAC operation, and  $E_{\text{AC}}$  is the energy cost per AC operation. We set  $E_{\text{MAC}} = 4.6\text{pJ}$  and  $E_{\text{AC}} = 0.9\text{pJ}$ , following (Zhou et al. 2023).

For ANNs, we calculate the FLOPs for each layer and multiply it by  $E_{\text{MAC}}$  to estimate the energy consumption.

For SNNs, when processing static RGB image data, the first layer is responsible for encoding the raw input into spike sequences. As a result, the operations in the first layer are computed as standard FLOPs<sup>(1)</sup>. The following layers, which process binary spike inputs, rely on synaptic operations for computation. We report the accuracy and the energy consumption for each prediction in different-scale ANNs and SNNs in Table 3.

As shown in Table 3, SNN models across all scales consistently exhibit lower energy consumption compared to their ANN counterparts. To provide a comprehensive evaluation, we first average the "Avg." Top-1 accuracy across the three partitioning settings (IID, Dir(0.5), and Dir(0.3)), and then compute the relative accuracy drop of SNNs compared to ANNs, alongside the average energy consumption. Overall, SNNs achieve a 43.74% reduction in power consumption with less than 1% accuracy degradation, which is an acceptable trade-off for energy-efficient learning. Interestingly, under the IID setting, our SFedHIFI even outperforms the ANN-based method in terms of accuracy. We attribute this to the advantage of the HIFI module in effectively leveraging heterogeneous information. However, as the Non-IID degree increases, the robustness advantage brought by the higher precision of ANNs becomes increasingly evident. Moreover, we observe that the benefits of SNNs are more pronounced in larger models: as network width increases, the energy savings of SNNs become more significant, while accuracy loss remains marginal. This energy-accuracy trade-off highlights the unique strength of SNNs in resource-constrained edge scenarios, demonstrating the practical potential of SFedHIFI in real-world SFL deployments.

## Conclusion

In this work, we propose SFedHIFI, a novel heterogeneous SFL framework. Based on Channel-wise Matrix Decomposition, SFedHIFI enables clients to adaptively deploy SNN models with varying widths according to their local resource constraints. To address the severe information loss caused by the inherent sparsity of SNNs, we design a Firing Rate-Based HIFI Module. Leveraging spiking dynamics, this module identifies high-information-density layers based on spike firing rates and performs Tucker information decoupling to fuse valuable heterogeneous layers from the model scale factors. Experimental results across multiple benchmarks demonstrate that SFedHIFI consistently outperforms all baseline SFL methods, effectively enabling heterogeneous federated learning via channel pruning. Ablation studies confirm the effectiveness of the proposed HIFI module and validate the mapping between firing rates and spiking information density. Furthermore, energy consumption analysis shows that SFedHIFI achieves notable energy savings with minimal accuracy trade-off. Overall, this work demonstrates the feasibility and effectiveness of constructing heterogeneous SFL systems, laying the foundation for broader deployment of neuromorphic intelligence in real-world federated learning scenarios.

## Acknowledgments

This research was supported by the NSFC (62376228), the Fundamental Research Funds for the Central Universities (JBK2406079), and the Graduate Representative Achievement Cultivation Project of Southwest University of Finance and Economics (JGS2024067).

## References

- Astrid, M.; and Lee, S.-I. 2017. Cp-decomposition with tensor power method for convolutional neural networks compression. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 115–118. IEEE.
- Balažević, I.; Allen, C.; and Hospedales, T. M. 2019. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*.
- Davies, M.; Srinivasa, N.; Lin, T.-H.; Chinya, G.; Cao, Y.; Choday, S. H.; Dimou, G.; Joshi, P.; Imam, N.; Jain, S.; et al. 2018. Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee Micro*, 38(1): 82–99.
- DeBole, M. V.; Taba, B.; Amir, A.; Akopyan, F.; Andreopoulos, A.; Risk, W. P.; Kusnitz, J.; Otero, C. O.; Nayak, T. K.; Appuswamy, R.; et al. 2019. TrueNorth: Accelerating from zero to 64 million neurons in 10 years. *Computer*, 52(5): 20–29.
- Deng, S.; Li, Y.; Zhang, S.; and Gu, S. 2022. Temporal Efficient Training of Spiking Neural Network via Gradient Reweighting. In *International Conference on Learning Representations*.
- Diao, E.; Ding, J.; and Tarokh, V. 2021. HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients. In *International Conference on Learning Representations*.
- Doutsis, E.; Fillatre, L.; Antonini, M.; and Tsakalides, P. 2021. Dynamic image quantization using leaky integrate-and-fire neurons. *IEEE Transactions on Image Processing*, 30: 4305–4315.
- Fang, W.; Chen, Y.; Ding, J.; Yu, Z.; Masquelier, T.; Chen, D.; Huang, L.; Zhou, H.; Li, G.; and Tian, Y. 2023. Spiking-Jelly: An open-source machine learning infrastructure platform for spike-based intelligence. *Science Advances*, 9(40): eadi1480.
- Gerstner, W.; Kistler, W. M.; Naud, R.; and Paninski, L. 2014. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.
- Haddadpour, F.; Kamani, M. M.; Mokhtari, A.; and Mahdavi, M. 2021. Federated learning with compression: Unified analysis and sharp guarantees. In *International Conference on Artificial Intelligence and Statistics*, 2350–2358. PMLR.
- Horvath, S.; Laskaridis, S.; Almeida, M.; Leontiadis, I.; Venieris, S.; and Lane, N. 2021. Fjord: Fair and accurate federated learning under heterogeneous targets with ordered dropout. *Advances in Neural Information Processing Systems*, 34: 12876–12889.
- Huang, W.; Ye, M.; Shi, Z.; Wan, G.; Li, H.; Du, B.; and Yang, Q. 2024. Federated learning for generalization, robustness, fairness: A survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Ji, Y.; Zhong, X.; Kou, Z.; Zhang, S.; Li, H.; and Yang, Y. 2023. Efficiency-boosting federated learning in wireless networks: A long-term perspective. *IEEE Transactions on Vehicular Technology*, 72(7): 9434–9447.
- Jia, Y.; Zhang, X.; Beheshti, A.; and Dou, W. 2024. Fedlps: Heterogeneous federated learning for multiple tasks with local parameter sharing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 12848–12856.
- Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, 5132–5143. PMLR.
- Kolda, T. G.; and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500.
- Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.
- Li, Q.; He, B.; and Song, D. 2021. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10713–10722.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2: 429–450.
- Maass, W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9): 1659–1671.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Mei, Y.; Guo, P.; Zhou, M.; and Patel, V. 2022. Resource-adaptive federated learning with all-in-one neural composition. *Advances in Neural Information Processing Systems*, 35: 4270–4284.
- Meng, L.; Qiao, G. C.; Zhang, X.; Bai, J.; Zuo, Y.; Zhou, P.; Liu, Y.; and Hu, S. G. 2023. An efficient pruning and fine-tuning method for deep spiking neural network. *Applied Intelligence*, 53: 1–14.
- Merolla, P. A.; Arthur, J. V.; Alvarez-Icaza, R.; Cassidy, A. S.; Sawada, J.; Akopyan, F.; Jackson, B. L.; Imam, N.; Guo, C.; Nakamura, Y.; et al. 2014. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197): 668–673.
- Nguyen, D. C.; Ding, M.; Pathirana, P. N.; Seneviratne, A.; Li, J.; and Poor, H. V. 2021. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 23(3): 1622–1658.
- Nguyen, M. V.; Zhao, L.; Deng, B.; Severa, W.; Xu, H.; and Wu, S. 2024. The Robustness of Spiking Neural Networks in Communication and its Application towards Network Efficiency in Federated Learning. In *2024 IEEE International Performance, Computing, and Communications Conference (IPCCC)*, 1–7. IEEE.

- Pan, Z.; Chua, Y.; Wu, J.; Zhang, M.; Li, H.; and Ambikairajah, E. 2020. An efficient and perceptually motivated auditory neural encoding and decoding algorithm for spiking neural networks. *Frontiers in neuroscience*, 13: 1420.
- Pei, J.; Deng, L.; Song, S.; Zhao, M.; Zhang, Y.; Wu, S.; Wang, G.; Zou, Z.; Wu, Z.; He, W.; et al. 2019. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature*, 572(7767): 106–111.
- Phan, A.-H.; Sobolev, K.; Sozykin, K.; Ermilov, D.; Gusak, J.; Tichavský, P.; Glukhov, V.; Oseledets, I.; and Cichocki, A. 2020. Stable low-rank tensor decomposition for compression of convolutional neural network. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, 522–539. Springer.
- Qiu, X.; Fernandez-Marques, J.; Gusmao, P. P.; Gao, Y.; Parcollet, T.; and Lane, N. D. 2022. ZeroFL: Efficient On-Device Training for Federated Learning with Local Sparsity. In *International Conference on Learning Representations*.
- Shah, S. M.; and Lau, V. K. 2021. Model compression for communication efficient federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9): 5937–5951.
- Shi, Y.; Nguyen, L.; Oh, S.; Liu, X.; and Kuzum, D. 2019. A soft-pruning method applied during training of spiking neural networks for in-memory computing applications. *Frontiers in neuroscience*, 13: 405.
- Singh, S.; Sarma, A.; Lu, S.; Sengupta, A.; Kandemir, M. T.; Neftci, E.; Narayanan, V.; and Das, C. R. 2022. Skipper: Enabling efficient snn training through activation-checkpointing and time-skipping. In *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 565–581. IEEE.
- Skatchkovsky, N.; Jang, H.; and Simeone, O. 2020. Federated neuromorphic learning of spiking neural networks for low-power edge intelligence. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8524–8528. IEEE.
- Tumpa, S. A.; Singh, S.; Khan, M. F. F.; Kandemir, M. T.; Narayanan, V.; and Das, C. R. 2023. Federated learning with spiking neural networks in heterogeneous systems. In *2023 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, 1–6. IEEE.
- Venkatesha, Y.; Kim, Y.; Tassioulas, L.; and Panda, P. 2021. Federated learning with spiking neural networks. *IEEE Transactions on Signal Processing*, 69: 6183–6194.
- Wang, J.; Liu, Q.; Liang, H.; Joshi, G.; and Poor, H. V. 2020. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33: 7611–7623.
- Wang, Y.; Duan, S.; and Chen, F. 2023. Efficient asynchronous federated neuromorphic learning of spiking neural networks. *Neurocomputing*, 557: 126686.
- Wu, C.; Wu, F.; Lyu, L.; Huang, Y.; and Xie, X. 2022. Communication-efficient federated learning via knowledge distillation. *Nature communications*, 13(1): 2032.
- Wu, J.; Xu, C.; Han, X.; Zhou, D.; Zhang, M.; Li, H.; and Tan, K. C. 2021. Progressive tandem learning for pattern recognition with deep spiking neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 7824–7840.
- Xiang, M.; Ioannidis, S.; Yeh, E.; Joe-Wong, C.; and Su, L. 2024. Efficient federated learning against heterogeneous and non-stationary client unavailability. *Advances in Neural Information Processing Systems*, 37: 104281–104328.
- Xiao, H.; Rasul, K.; and Vollgraf, R. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.
- Xie, X.; Feng, J.; Liu, G.; Zhan, Q.; Liu, Z.; and Zhang, M. 2024. Federated learning for spiking neural networks by hint-layer knowledge distillation. *Applied Soft Computing*, 163: 111901.
- Xu, J.; and Wang, H. 2020. Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective. *IEEE Transactions on Wireless Communications*, 20(2): 1188–1200.
- Xu, Y.; Liao, Y.; Xu, H.; Ma, Z.; Wang, L.; and Liu, J. 2022. Adaptive control of local updating and model compression for efficient federated learning. *IEEE Transactions on Mobile Computing*, 22(10): 5675–5689.
- Yang, H.; Lam, K.-Y.; Xiao, L.; Xiong, Z.; Hu, H.; Niyato, D.; and Vincent Poor, H. 2022. Lead federated neuromorphic learning for wireless edge artificial intelligence. *Nature communications*, 13(1): 4269.
- Yu, Y.; Yan, Y.; Cai, J.; and Jin, Y. 2024. Heterogeneous federated learning with convolutional and spiking neural networks. In *International Workshop on Trustworthy Federated Learning*, 93–105. Springer.
- Zhou, Z.; Zhu, Y.; He, C.; Wang, Y.; YAN, S.; Tian, Y.; and Yuan, L. 2023. Spikformer: When Spiking Neural Network Meets Transformer. In *The Eleventh International Conference on Learning Representations*.